基于 DCNN 分类的图像相关度度量 *

王会勇，谢春杰，张晓明，孙晓领

(河北科技大学 信息科学与工程学院, 石家庄 050018)

摘　要：在衡量图像之间的相关度时，图像的物理特征（颜色分布、灰度值等）所能表达的内容可能并非十分全面，因此有必要参考图像视觉所包含的语义信息衡量图像之间的相关度。为此提出了一种基于深度卷积神经网络（deep convolutional neural networks）分类模型的度量图像相关度的方法，利用模型为图像绑定来自于 WordNet 的语义标签，并参照 WordNet 结构对标签进行过滤和扩展，利用概念集合计算图像相关度。与人工判定的样本数据比较，Pearson 相关系数峰值能够达到 0.73，证明该方法在衡量图像相关度时具有一定的效果。

关键词：相关度；深度卷积神经网络；WordNet；过滤；扩展

中图分类号：TP391　　doi: 10.19734/j.issn.1001-3695.2018.04.0487

Image relativity metric based on classification with DCNN

Wang Huiyong, Xie Chunjie, Zhang Xiaoming, Sun Xiaoling

(*School of Information Science & Engineering*, *Hebei University of Science & Technology*, *Shijiazhuang* 050018, *China*)

Abstract: When measuring the similarity between images, the content of the physical features (Color Layout Descriptor, Gray Histogram Descriptor, etc.) may not be very comprehensive, so it is necessary to refer to the semantic information contained in image vision to measure the relativity between images. In this paper, we propose a method based on Deep Convolutional Neural Networks classification model to measure image correlation. The model is used to bind the semantic label from WordNet, and the label is filter and expand according to WordNet structure, and the concept set is used to calculate image relativity. Compared with the manually determined sample data, the peak value of Pearson correlation coefficient can reach 0.73, which proves that this method has a certain effect in the measurement of image correlation.

Key words: relativity; DCNN; WordNet; filter; expand

0　引言

随着深度学习方法在计算机视觉领域取得了突破性的进展，传统的图像分类问题已经得到了相对良好的解决，跨模态数据（文本数据和多媒体数据）的深度融合逐渐成为研究的热点，例如图片标题生成[1]、文本生成图像[2~4]等。在另外一些研究任务中，如构建大型多模态知识图谱[5]，或者跨模态的实体链接[6]等,也都将跨模态数据的融合视为关键的核心技术问题。跨模态数据的融合，能够得到图像视觉语义在文本语义空间的表达，利用文本语义空间的抽象层次结构,可以实现语义内容的扩展。能否将跨模态数据融合的结果应用于图像相关度的计算，是一个值得研究的问题。

本文将基于深度卷积神经网络模型的图像分类视为跨模态数据的浅层融合，并且将采用多模态表示学习方法[7]实现的跨模态数据融合视为深度融合。相较而言，二者都需要处理图像的高级特征，不同之处在于后者同时融合了文本特征的处理。由于不需要考虑提取文本特征以及将图像特征和文本特征联合嵌入到共享空间[8]中的问题，因此基于深度卷积神经网络模型的分类方法所需的成本相对不高。

在历届 ImageNet[9,10]大规模视觉识别挑战赛（ILSVRC）中，先后出现了很多经典的深度神经网络模型，例如 Alexnet[11]、VGG[12]、deep residual learning[13]等。在最新的图像分类比赛任务中，top-5 的错误率已经降到了 0.0901，这样的准确性已经达到了很高的水平。结合方法成本和模型的效果，本文设计了基于深度卷积神经网络图像分类的图像相关度计算方法。

当利用 ImageNet 数据集对深度卷积神经网络模型进行训练后，模型能够为图像绑定来源于 WordNet[14]的资源标签。WordNet 可以视为一个较为权威的文本语义空间，参照

收稿日期：2018-04-18；**修回日期**：2018-05-30　**基金项目**：河北省自然科学基金资助项目（F2018208116）；河北省科技计划项目（16210312D）；河北省教育厅资助科研项目（ZD2015099）

作者简介：王会勇（1980-），男，河北石家庄人，讲师，博士研究生，主要研究方向为模式识别与机器学习、语义 Web、知识图谱（wanghuiyong815@163.com）；谢春杰（1992-），男，硕士研究生，主要研究方向为语义 Web、知识图谱；张晓明（1975-），男，教授，博士研究生，主要研究方向为语义 Web、知识图谱；孙晓领（1994-），女，硕士研究生，主要研究方向为语义 Web、知识图谱.

WordNet 定义的类型结构，可以设计规则对概念进行过滤和扩展，从而实现语义上下文的获取。扩展得到的概念，视为与图像所包含的语义信息存在关联。

本文基于上述思想，完成了方法的设计。包括相关的扩展规则，可调整的参数以及计算方法。首先对图像分类标签进行适当的过滤，然后按照扩展规则，生成得到图像关联的概念集合。概念集合中的元素不是彼此独立的，而是存在按照语义划分的上下位关系。概念间的上下位关系对衡量相关性起到了一定的决定影响，因此提出了一种将上下位关系转换为权重的规则，利用图像关联的概念集合设计了计算图像与图像之间相关度的方法。

1　相关工作

本文提到的相关度与相似度没有明确的界限，由于本文设计的方法是将图像融合的文本所在的语义空间的扩展内容作为衡量的标准，因此使用相关度描述更加贴切。在衡量图像之间的关系时，需要对图像所表达的内容进行分析。按照使用的手段不同，可以将目前衡量图像相关性的方法分为基于物理特征的方法和深度学习方法两类：

基于物理特征的方法。这类方法主要对图像的物理特征进行分析，例如灰度直方图（gray histogram descriptor）、方向梯度直方图（histogram of oriented gradients descriptor）和颜色布局（color layout descriptor）等。Ferrada 等人[15]在构建大型的多模态知识图谱时，依据上述特征对图像之间的相似度进行了评价，并在此基础上设计了相关的查询方法。事实上图像的物理特征往往并不能将图像所包含的语义内容全部表达出来，因此这类方法具有一定的局限性。

基于深度学习的方法。这类方法主要利用神经网络模型对输入的图像进行卷积计算，得到图像的特征向量后，直接利用特征向量进行相似度计算。相对于第一类方法，基于深度学习的方法能够得到更加理想的效果。例如 Chopra 等人[16]将 Siamese[17]网络应用于人脸识别任务中，取得了较好的结果。Zagoruyko 等人[18]对 Siamese 网络进行了改进，并融合了空间金字塔池化采样[19]方法，进一步提升了模型的精度。本文认为，图像所包含的语义信息是具有扩展性的，如图 1 所示，当直接利用图像 a 和图像 b 的特征向量进行计算时，往往不能捕捉到两个图像内容存在的隐含的联系。

当人类观察者在观察图 1 中的两个图像 a 和 b 时，会依据个人具有的背景知识对图像中的内容进行分辨，并产生相关的联想。例如图像 a 中的俄罗斯蓝猫属于猫类，猫类又属于哺乳动物。图像 b 中的印度辊属于鸟类。而哺乳动物和鸟类都属于动物。这样的联想能够使得两个图像在文本语义空间中取得一定的关联性。在一些实际情况中，观察者可能无法准确判断图 1 中猫的种类和鸟的种类，甚至可能发生错误的判断。但是这样的情况对于衡量两个图像之间的相关度没有太大影响。

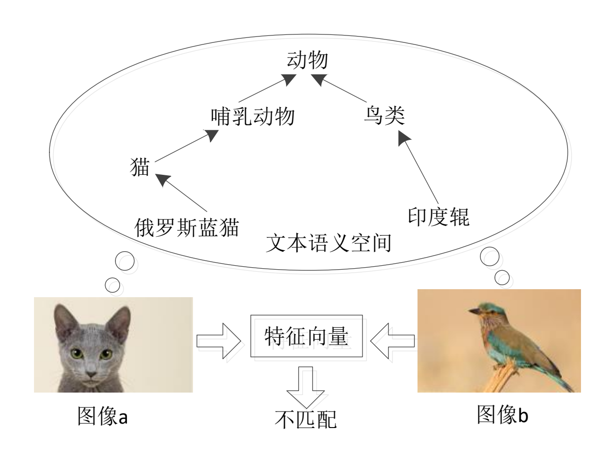


图 1　图像的内在关联

Fig.1　internal relation of images

本文利用深度卷积神经网络模型为图像绑定分类标签，并依据 WordNet 所提供的较为权威的内容扩展上下位关系，以达到挖掘图像之间隐含关联的目的。模型训练过程中使用到了 ImageNet 数据集，一个大规模图像语料库。ImageNet 根据 WordNet 层次结构组织图像数据，为每一个有意义的概念（synset）提供了 1000 张图像说明，图像的质量控制和注释都是由人工完成的。最新的 ImageNet 数据集为 21841 个概念提供了共 14197122 张图像，训练好的模型可以预测 1000 个类别。

WordNet 是一个由普林斯顿大学发起的研究项目，是一个按照词语的词义组织词汇信息的大型在线英文分类数据库系统。每一个词语包含一个或者多个概念，每个概念都对应着一个或者多个单词及单词短语的描述，成为同义词集合（synset）。目前，WordNet3.0 中包含了 10 万个同义词，其中 80%以上为名词，其定义的上下位层次总数为 19 层。

2　问题定义

利用图像视觉内容关联的 WordNet 概念计算图像之间的相关度是本文着重解决的最核心的问题。在设计计算方法时，文本将与图像视觉内容不相符的 WordNet 概念（分类结果噪声）以及抽象层次过高的 WordNet 概念（概念扩展噪声）视为计算过程中的噪声。设计有效的方法解决分类结果的噪声问题和概念扩展的噪声问题是解决图像相关度计算问题的必要基础。

2.1　分类结果的噪声问题

本文在获取图像与文本的融合结果时，采用了与图像分类相同的方法，但是对得到的结果采取了不同的处理策略。在图像分类任务中，期望得到的是准确的图像从属的类别。而在本文相关的任务中，则是将模型得到的标签视为图像视觉语义信息在文本空间的表达。如图 2 所示，对于一张图像的分类结果，如果直接进行标签对比，判断得到的结果可能是分类错误。例如图 2 中的鸟为印度辊，但是分类的结果并不包含正确标签，因此判断分类结果为错误。而如果将这些标签作为图像视觉语义信息在文本空间的表达，当在一个具有抽象层次的结构中审视这些标签时，可能这些标签对于分析图像所表达的内容是有帮助的。即没有得到对图像的最为精确的描述，但是可以推出

图中描述的是和鸟相关的内容。

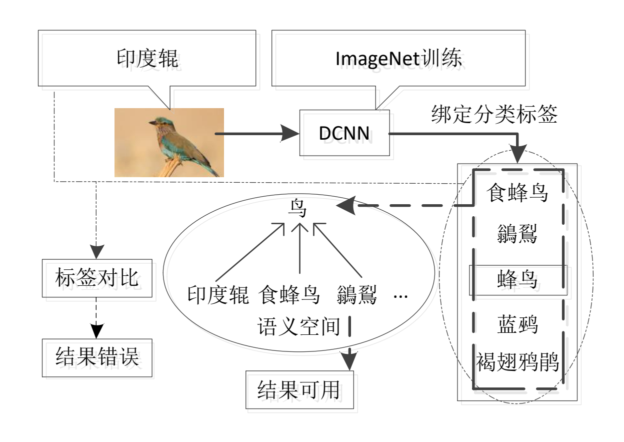


图 2　与图像分类任务的不同

Fig.2　Different from image classification

ImageNet 在对模型分类结果的准确率进行测评时，采用了统计 top-5 准确率的方法，即模型为图像分配的 5 个标签中只要有一个是准确的，结果就认定为准确。这样的结果存在的问题是，5 个标签的内容不全都是准确的（分类标签与图像视觉内容不相符）。由于本文利用训练好的深度神经网络模型对图像进行分类时，也保留模型为每个图像分配的分类标签的 top-5，因此这样的结果或多或少都包含了一定的噪声，对图像相关度的计算结果产生影响。为了减少这样的噪声，本文设计了基于有效距离的筛选方法，尽可能过滤分类结果中存在的与图像视觉内容不相符的标签。

2.2　概念扩展的噪声问题

由于分类模型为图像的绑定的分类标签对应着 WordNet 中的某一个确切的概念，因此依据 WordNet 定义的类型结构，可以获得每个概念的上位词，概念的上位词还能继续向上扩展，最后都链接到 entity 概念。越上层的概念意味着抽象层次越高。当某一个概念向上扩展至 entity 所在的抽象层次，再将相关概念作为衡量图像相关度的依据时，由于抽象层次过高（所有的概念都能关联到 entity 概念），得到的结果将变得不可靠。

另一方面，分类模型为每个图像分配多个分类标签，每个概念向上扩展的层次可能很多，因此不加限制的扩展将会得到一个庞大的概念集合。如果不考虑抽象层次过高带来的噪声问题，在计算相关度时，一些不相关的概念将会影响结果的准确性。因此为了尽量避免出现这样的问题，本文设计了限制扩展层次的方法对扩展结果进行有效控制。

2.3　图像相关度的计算问题

在处理完分类结果的噪声与概念扩展的噪声之后，本文利用图像关联的 WordNet 概念计算图像之间的相关度。与基于 WordNet 的词语相似度计算不同的是，在本文设计的计算图像相关度方法中，需要同时考虑的是一组概念，而不是两个单独的概念。因此不能采用相同的方法进行计算，因为如果逐一计算两个概念间的相似度，需要以牺牲时间效率为代价，并且对计算得到的多个结果进行合理的融合也是一个较难的问题。针对相关度的计算问题，本文设计了一种提取有效概念集合的方法，并基于 Jaccard 相似度计算方法设计了计算图像相关度的方法。本文还提出了一种计算权值的方法，用于将概念间的上下位关系融入到图像相关度的计算中。

3　方法设计与实现

针对第 2 章提出的三个问题，分别设计了不同的解决方案，并最终整合为一个完整的模型。模型主要包含两个模块，如图 3 所示。两个模块分别为分类标签过滤模块和相关度计算模块。模型输入为两张图像，输出为一个位于 0~1 的数值。对输入的处理流程分为以下几个步骤：

a)利用 DCNN 分类器分别为输入图像绑定分类标签，每个图像会绑定 5 个标签。

b)利用基于有效距离的过滤模块对标签进行适当的过滤。过滤得到有效分类标签集合，它们将被作为下一个模块的输入。

c)依据 WordNet 结构生成相关的概念集合，利用集合计算相关度的值，输出一个 0 到 1 之间的数值。

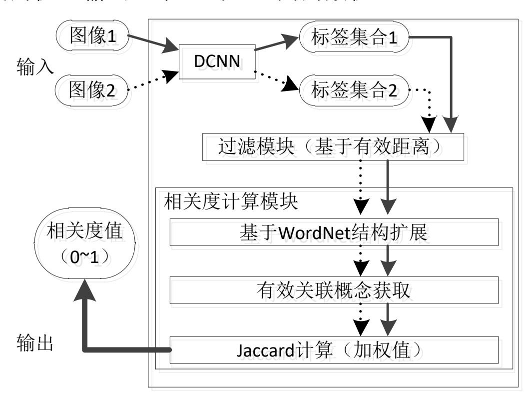


图 3　方法流程图

Fig.3　Flow diagram

过滤模块主要用于降低分类结果中出现的噪声，有一个人工调节的参数，即阈值 r，用于限定有效距离排名的结果，从而达到过滤的目的。相关度计算模块有一个人工调节的参数 t，用于限制向上扩展的层次。在图 3 中，两种不同风格的细线条表示对两个图像的处理流程，下面展开详细说明。

3.1　基于有效距离的分类标签筛选

在 WordNet 定义的类型结构树中，由一个概念到达另一个不同的概念所经历的最短路径称长度为两个概念间的最短距离。通过深度神经网络模型对图像进行分类处理，得到图像对应的 5 个分类标签，对于任意一个分类标签，它和其余 4 个分类标签的最短距离的平均值，称为有效距离。对于每一个分类标签，都根据式（1）所示的有效值计算公式确定其对应的有效值。

$$effe(c_i,C)=\begin{cases}0 & aPath(c_i,C)\ is\ not\ top-r\\ 1 & aPath(c_i,C)\ is\ top-r\end{cases} \tag{1}$$

其中：$effe(c_1,C)$表示分类标签 c_1 对应的有效值，只有 0 或 1 两种情况。$aPath(c1,C)$表示分类标签 c_1 在集合 C 中的有效距离，集合 C 为 c_1 所在的 5 个分类标签组成的集合。r 表示人工可调节的阈值参数，当前元素的有效距离若在当前集合所有元素的有效距离的升序排名属于 $top\text{-}r$，则其有效值为 1，r 的取值在

1~5，当 r 值为 5 时，表示不进行过滤。得到每个分类标签的有效值后，根据有效值对分类标签进行筛选，筛选后的有效分类标签集合定义为 S，表示方式由式（2）所示。

$$S=\{c|c\in C\wedge effe(c,C)=1\} \tag{2}$$

其中：$effe(c,C)$按照式（1）所示的计算方法进行计算。如图 4 所示，对基于有效距离的分类标签筛选策略进行了更直观的说明。DCNN 模型为图像分配五个标签，根据 WordNet 结构计算标签之间的距离得到距离矩阵，计算任意一个标签与剩余标签之间的平均距离，并按照平均距离大小对标签进行排名，最终根据设置的 r 值对排名结果进行筛选，构建有效分类标签集合 S。

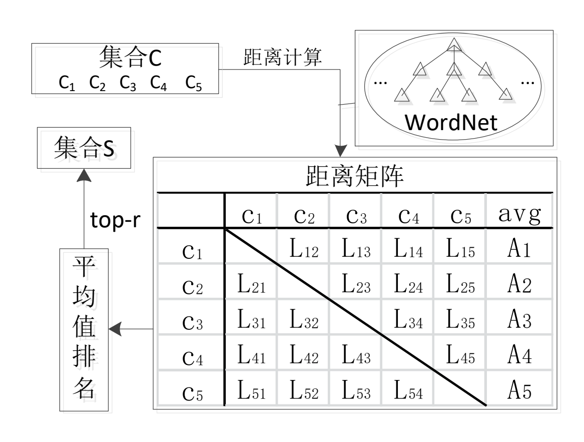


图 4　基于有效距离的分类标签过滤方法流程

Fig.4　Method flow of classified label filtering based on effective distance

3.2　基于上位词扩展集合的图像相关度计算

假设有两张图像分别为图像 A 和图像 B，S_A 表示图像 A 对应的有效分类标签集合，S_B 表示图像 B 对应的有效分类标签集合。 E_A 表示图像 A 对应的扩展集合，E_B 表示图像 B 对应的扩展集合。E_A 和 E_B 满足式（3）和（4）所示的条件。

$$E_A=ect(S_A,t) \tag{3}$$

$$E_B=ect(S_B,t) \tag{4}$$

在 E_A 和 E_B 的基础上提出了有效关联集合的概念。有效关联集合是一个相对的概念，当在对比两个图像各自的扩展集合时，按照以下两个判断条件生成有效关联集合：

a) 当前元素与另一个集合中的某个元素相同，则当前元素及其所在集合中属于当前元素的下位词的元素，均属于有效关联集合。

b) 当前元素在另一个集合中不存在相同元素，则当前元素属于有效关联集合。

有效关联集合用 O 表示，O_{AB} 表示图像 A 相对于图像 B 的有效关联集合，O_{BA} 表示图像 B 相对与图像 A 的有效关联集合。假设 t 和 r 的值都为 3，按照上述两个判断关联集合的示意图如图 5 所示。

不难发现，有效关联集合是一个相对概念，图像 A 对于不同的图像 B 和图像 C，得到的有效关联集合不一定是相同的。在图 5 中，不同的花纹的形状代表不同的概念，相同花纹的形状代表相同的概念。其中空心圆和空心十二边形是满足两个判断条件中的条件 b 的，其余概念是满足两个判断条件中的条件 a 的。

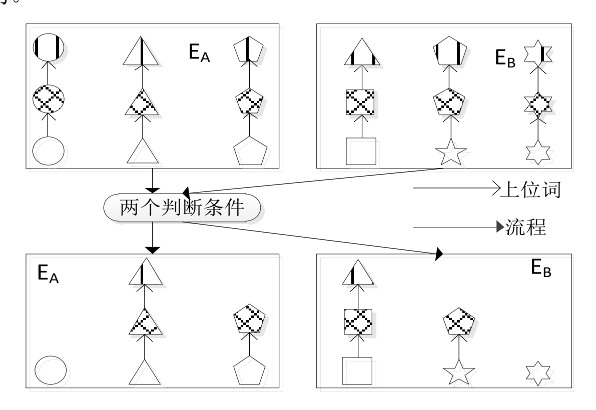


图 5　有效关联集合生成示意图

Fig.5　Generate schematics of associated sets

得到图像的有效关联概念集合后，就能够计算两个图像的相关度。本文设计的相关度计算公式是以 Jaccard 计算公式为基础设计的。Jaccard 计算公式如式（5）所示。

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|}=\frac{|A\cap B|}{|A|+|B|-|A\cap B|} \tag{5}$$

在设计相关度计算公式时，首先确定了 Jaccard 公式的两个输入矩阵。将有效关联集合和扩展集合作为计算的内容，按照式（6）所示的相关度计算函数计算。

$$corr(A,B)=\frac{|O_{AB}\cup O_{BA}|}{|E_A\cup E_B|} \tag{6}$$

扩展概念集合中的元素在 WordNet 定义的体系结构中，存在上下位关系。按照式（6）的计算方法，每个元素所占的权重均为 1。在单独对比两个 WordNet 概念间时，深度越小的节点对概念间相似关系的影响越大[20]。在对比一组概念时，本文同样遵循这样的原则。并提出了一种简单的计算权重的方法。对集合 E_A 中的元素，其权值按照式（7）所示的方法计算。

$$pow(c)=\frac{t_0}{1+t} \tag{7}$$

其中：t 表示设定好的扩展层次，t_0 表示当前概念的相对距离，若元素为起始扩展概念（属于集合 S），则其相对距离为 1。

计算出每个概念的权重后，在式（6）的基础上设计了基于权重的相关度计算函数，如式（8）所示。

$$corrP(A,B)=\frac{\sum_{i\in(O_{AB}\cup O_{BA})}pow(i)}{\sum_{j\in(E_A\cup E_B)}pow(j)} \tag{8}$$

其中：pow(i)和 pow(j)按照式（7）的方法计算。将概念之间的层次关系作为影响计算的权重，能够提升计算结果的效果，在实验部分证实了这样的猜想。

4　实验及结果

4.1　采集人工评价数据

本文设置了十组样本图像对作为相关度计算效果评价的测试集，并组织了 50 个观察者对图像的相关度进行评分，分数最低分为 0 分，最高分为 4 分。在对图像的相关度进行评分时，

观察者需要考虑两个图像所蕴含的语义内容，并根据个人的判断对图像的相关度进行评价。对 10 组样本图像对进行评价的统计结果如表 1 所示。

在表 1 中，按照评分值的升序顺序对样本数据进行了展示。其中最低分为 0.30 分，最高分为 3.36 分。按照评分的升序序列对十组图像对进行了编号，评分为 0.30 的图像对编号为第一组，评分为 0.46 的图相对编号为第二组，以此类推。人工评价数据最终将作为参照，在后续的实验中用于评价本文方法的结果。

表 1　10 组样本图像对的人工评价

Table 1　Artificial evaluation of 10 groups of sample images

图像对		评分	图像对		评分
		0.30			1.73
		0.46			1.91
		1.10			1.98
		1.18			2.18
		1.46			3.36

4.2　r 值的确定

本文选择的用于图像分类的深度神经网络模型为 caffeNet，一个改进后的 AlexNet 模型，将 ILSVRC-2012 数据集作为图像分类的训练集。通过基于 ILSVRC-2012 数据集的 50000 次迭代训练之后，利用 ILSVRC-2012 测试集的评测结果显示 top-1 的错误率为 7%。

本文从维基百科中随机挑选了 200 张图像作为确定 r 值的测试数据集，由于篇幅有限，在此不做展示。首先利用训练好的 caffeNet 模型对 200 张图形进行分类处理，每个图像保留准确率排名为 top-5 的分类标签。然后对于每张图像所对应的分类标签按照式（2）所示的方法生成有效分类标签集合。最后通过设置不同的 r 值对分类标签进行过滤，通过对比不同 r 值情况下过滤的准确性，确定 r 值的合理取值。

在不考虑重复的情况下，通过对 200 张图像处理，能够得到 1000 个分类标签。对于不同的 r 值，本文通过统计过滤的准确度来评判当前 r 值的效果，统计结果如图 6 所示。过滤结果的准确度为过滤标签中应该被过滤的数量与过滤标签总数的比值。在图 6 中，横轴表示不同的 r 值，纵轴表示当前 r 值情况下过滤的准确度。通过测试结果不难看出，随着 r 值的增加，过滤结果总体呈现波动趋势。当 r 值为 3 时，过滤的准确度达到峰值 0.74，表明此时的过滤效果最好。

由于分类模型为每张图像分配了 5 个标签，因此 r 的有效取值范围时[1,4]，r 取整数。r 值越大，有效标签集合的元素越多；反之，有效集合标签的元素越少。由于图像的视觉特征是固定的，尽管分类模型存在误差，但是为图像分配的标签在语义层面应该是互相接近的。因此按照 r 值的过滤遵循了这个规律，如果几个概念彼此间的距离都很近，则认为这几个概念与图像视觉内容的相关度更高。当 r 值偏大时，这种有效距离的范围也变大，因此过滤效果不明显导致准确度下降。当 r 值过小时，相对准确的标签可能会被过滤掉，因为不能保证准确概念与其他概念间的有效距离始终是最短的。

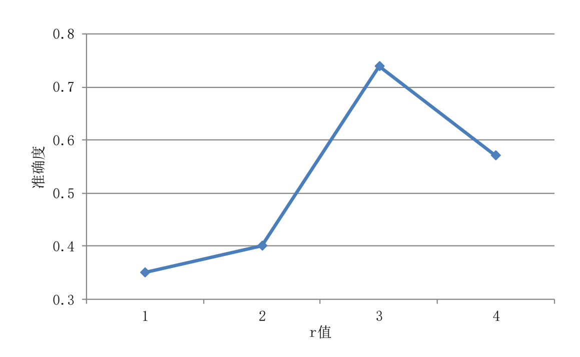


图 6　利用 200 张图像测试的不同 r 值的过滤准确度

Fig.6　Filtering accuracy of different *r* tested with 200 images

4.3　相关度的计算评价及结果

在确定 r 值之后，就可以得到图像的扩展集合，扩展时使用的数据集为 WordNet 3.0。利用本文设计的方法实现了图像相关度的计算，并与表 1 中的数据进行了对比。通过计算表 1 所示的图像样本的相关度，最终和人工给出的评分进行比较。实验中设置了不同的 r 值和 t 值，并对比了在最优 r 值情况下加权 Jaccard 计算与不加权 Jaccard 计算方法得到结果的变化情况。将依据本文方法计算的结果与人工评价结果之间的 pearson 系数作为参考指标，最终对比结果如图 7 所示。

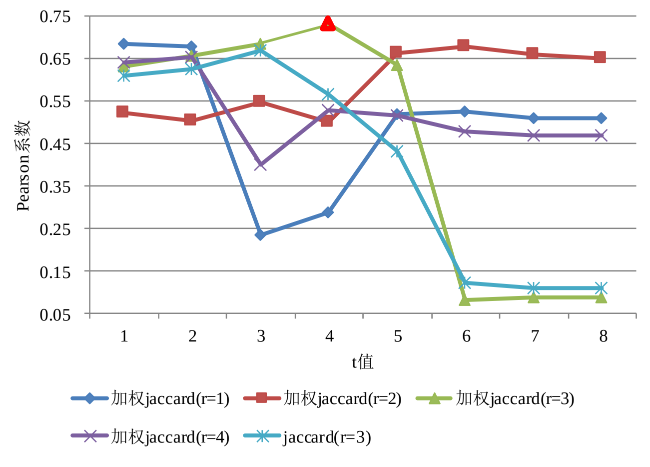


图 7　图像相关度计算结果测评

Fig.7　Calculation results of image correlation

在图 7 中，纵轴代表 pearson 系数，横轴代表不同的 t 值。通过设置不同的 r 值和 t 值，计算的效果呈现波动趋势。当 Pearson 系数达到峰值 0.730 时，r 值和 t 值分别为 3 和 4。最终结果显示，最优 r 值与前述实验得出的结论一致，即 r 的取值变化对计算效果有影响，并且在当前实验中，最优 r 值为 3，说明对模型分类噪声进行的处理是有效的。

另一方面，t 值的选取对计算的效果也是有影响的。t 值越大，表明扩展得到的概念层次越高；反之，扩展得到的概念抽

象层次越低。当 t 值取值过小时，扩展得到的概念相对较少，因此能够作为桥梁的公共上位词数量也较少，导致参与计算的概念集合之间相关性不高，计算效果不理想。当 t 值过大时，扩展得到的概念增多，公共上位词数量过多，导致参与计算的概念集合在较高的抽象层次取得关联，计算效果也不理想。WordNet 中常见的名词层次数量很少有超过 10 层的，因此当 t 值达到某一个区间后，扩展概念集合和有效关联概念集合变动不大，整体的计算效果趋于平稳。

在 r 值固定的情况下对比了加权计算与不加权计算的结果，数据表明加权计算的效果更好。将概念间的上下位关系转换为权值后，概念在层次结构中的位置对结果也会产生影响。公共上位词的权值最大，按照层次结构向下依次递减。当不存在公共上位词时，有效关联概念集合的元素所占权重很小，相对于不加权的计算方法，这样得到的结果更小。最终的实验结果表明，r 值和 t 值共同影响着计算结果。r 值可以通过基于有效距离的标签过滤实验方法确定，t 值可以根据最终的结果进行人工调整。同时，加权计算得到的结果优于不加权计算得到的结果。

在表 2 中对参数 r 设置为 3、参数 t 设置为 4 的情况下计算的加权计算结果进行了展示。由于人工评价采用了 4 分制，而计算的结果归一化为[0,1]的数值，因此将人工评价的结果也进行了归一化处理，在原始评分的基础上，缩小 4 倍得到一个归一化的结果，然后再作为计算结果的参照。

通过表 2 的展示可以发现，在当前参数设置情况下，计算的结果与人工评价的结果相差不大。其中，对于人工评价分值不高的图像对，计算结果基本吻合。对于人工评价分值高的图相对，计算结果出现了较低的情况。在第 5 节对实验的情况进行了进一步的分析，并提出了未来的工作。

表 2　加权计算结果

Table 2　Weighted calculation results

图像对	人工	加权计算	图像对	人工	加权计算
第一组	0.08	0.06	第六组	0.43	0.11
第二组	0.12	0.09	第七组	0.48	0.43
第三组	0.28	0.08	第八组	0.50	0.31
第四组	0.30	0.34	第九组	0.55	0.13
第五组	0.37	0.45	第十组	0.84	0.71

5　结束语

通过与人工评价的结果进行对比，可以看出本文设计的基于 DCNN 分类的图像相关度度量方法可以取得一定的效果。该方法能够挖掘图像之间隐含的关联关系，并依据 WordNet 定义的结构进行上下文扩展，进一步的将这种关系转换为数值形式。对于人工评价高的图像对，出现计算结果较低的情况，本文认为是由于得到的分类标签在 WordNet 中距离较近，由于扩展层次参数 t 的限制，会导致扩展集合中的元素数量变大，而有效关联集合中的元素数量不变，造成了结果偏小的情况。另外在实现过程中发现了该方法存在的一些其他问题，主要包括两方面，一方面是图像分类模型的准确度问题，另一方面是计算方法的合理性问题。针对这两方面的问题提出了未来的工作：

a）更多第三方知识库的选择。针对计算结果的误差问题，考虑集合更多的第三方知识库，如 ConceptNet[21]等，对图像关联的文本概念进行扩展，并作为计算的内容。

b）更加优化的深度卷积神经网络模型。针对分类模型的准确性问题，考虑用精度更高的模型代替，同时引入基于图像的实体识别方法，对图像的内容进行更深层次的挖掘，而不仅限于图像的分类。

c）计算方法的优化。在加权值的 Jaccard 计算方法的基础上，考虑进一步的改进权值的计算方法，保证计算结果的最优化。

此外，在构建多媒体知识图谱时，可以利用本文的思想，结合图像与知识图谱中的资源存在的关联关系，以及知识图谱中资源与资源之间的关联关系，确定图像与图像之间的关联关系。

参考文献：

[1] Vinyals O, Toshev A, Bengio S, *et al.* Show and tell: a neural image caption generator [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 3156-3164.

[2] Reed S, Akata Z, Yan Xinchen, *et al.* Generative adversarial text to image synthesis [C]// Proc of the 33rd International Conference on Machine Learning. New York: ACM Press, 2016: 1060-1069.

[3] Liu Pengfei, Qiu Xipeng, Huang Xuanjing. Adversarial multi-task learning for text classification [C]// Proc of Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2017: 1-10.

[4] Mansimov E, Parisotto E, Ba J L, *et al.* Generating images from captions with attention [C]/ Proc of International Conference on Learning Representations. (2016-02-04) . https://arxiv. org/pdf/1511. 02793. pdf.

[5] Zhu Yuke, Zhang Ce, Ré C, *et al.* Building a large-scale multimodal knowledge base system for answering visual queries [EB/OL]. arXiv preprint arXiv: 1507. 05670, 2015. https://arxiv. org/pdf/1507. 05670. pdf.

[6] Venkitasubramanian A N, Tuytelaars T, Moens M F. Entity linking across vision and language [J/OL]. Multimedia Tools and Applications, 2017. http://doi. org/10. 1007/s11042-017-4732-8.

[7] Ngiam J, Khosla A, Kim M, *et al.* Multimodal deep learning [C]// Proc of International Conference on Machine Learning. 2011: 689-696.

[8] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines [C]// Proc of the 27th International Conference on Machine Learning. 2010: 807-814.

[9] Deng Jia, Dong Wei, Socher R, *et al.* ImageNet: a large-scale hierarchical image database [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2009: 248-255.

[10] Russakovsky O, Deng Jia, Su Hao, *et al.* Imagenet large scale visual



recognition challenge [EB/OL]. International Journal of Computer Vision, 2015. https://arxiv. org/pdf/1409. 0575. pdf.

[11] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems. 2012: 1097-1105.

[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. arXiv preprint arXiv: 1409. 1556, 2014. https://arxiv. org/pdf/1409. 1556. pdf.

[13] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[14] Atkinson J, Ferreira A, Aravena E, *et al.* Discovering implicit intention-level knowledge from natural-language texts [J/OL]. Knowledge Based Systems, 2009. https://doi. org/10. 1007/978-1-84882-171-2_18.

[15] Ferrada S, Bustos B, Hogan A. IMGpedia: a linked dataset with content-based analysis of wikimedia images [C]// Proc of International Semantic Web Conference. Cham: Springer, 2017: 84-93.

[16] Chopra S, Hadsell R, Lecun Y. Learning a similarity metric discriminatively, with application to face verification [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2005: 539-546.

[17] Bromley J, Guyon I, Lecun Y, *et al.* Signature verification using a"Siamese"time delay neural network [C]// Proc of International Conference on Neural Information Processing Systems. San Franciscio: Morgan Kaufmann Publishers Inc. 1993: 737-744.

[18] Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4353-4361.

[19] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [C]// Provc of European Conference on Computer Vision. Cham: Springer, 2014: 346-361.

[20] Ahsaee M G, Naghibzadeh M, Naeini S E Y. Semantic similarity assessment of words using weighted WordNet [J/OL]. International Journal of Machine Learning and Cybernetics, 2014. https://doi. org/10. 1007/s13042-012-0135-3.

[21] Speer R, Havasi C. Representing General Relational Knowledge in ConceptNet 5 [C]// Language Resources and Evaluation, 2012: 3679-3686.